REMARKS

Claims 1-8, 10, 12-25, 27, and 29-36 are pending.

The final Office Action mailed May 20, 2004 allowed claims 7-8 and 24-25, objected to claims 15 and 32 as allowable but dependent on a rejected base claims, and rejected claims 17 and 34-36 under 35 U.S.C. § 102 as anticipated by the *Background* section; claims 1-3 and 18-20 as obvious under 35 U.S.C. § 103 based on *Rastogi et al.* (U.S. 6,247,016) in view of *Shimoji et al.* ("Data Clustering with Entropical Scheduling"); claims 4 and 21 over *Rastogi et al.*, *Shimoji et al.*, and *Background*; claims 5 and 22 over *Rastogi et al.*, *Shimoji et al.*, *Background*, and *Hall et al.* ("Generating Fuzzy Rules from Data"); claims 6 and 23 over *Rastogi et al.*, *Shimoji et al.*, and *Shafer et al.* ("SPRINT: A Scalable Parallel Classifier for Data Mining" 1996); claims 1-5 and 18-22 over *Janikow* ("Fuzzy Decision Trees: Issues and Methods") and *Choe et al.* ("On the Optimal Choice of Parameters in a Fuzzy C-Means Algorithm"); claims 6 and 23 over *Janikow*, *Choe et al.*, and *Shafer et al.*; claims 10, 12, 16, 27, 29, and 33 over *Janikow*; claims 13 and 30 over *Janikow* and *Choe et al.*; and claims 14 and 31 over *Janikow* and *Shafer et al.*

The rejection of claims 1-3 and 18-22 based on *Rastogi et al.* in view of *Shimoji et al.* is respectfully traversed because *Rastogi et al.* in view of *Shimoji et al.* fail to disclose the limitations of these claims. For example, independent claims 1 and 18 recite: "performing a cluster analysis along the selected feature to group the data into one or more clusters based on distances between the data and respective one or more centers of the one or more clusters."

This limitation is not shown in *Rastogi et al.* Rather, *Rastogi et al.* is directed to a decision tree classifier with integrated building and pruning phases (Title). *Rastogi et al.* involves sample records having multiple attributes, the sample records being identified or "tagged" with a special classifying attribute which indicates a **class** to which the record belongs. For example, as shown in FIG. 1, a training set has sample records identifying the salary level

(continuous attributes) and education level (categorical attributes) of a group of applicants for loan approval. Each record is tagged with either an "accept" classifying attribute or a "reject" classifying attribute, depending upon the parameters for acceptance or rejection set by the user of the database (col. 2:33-49). *Rastogi et al.* discloses that its "tree is built breadth-first by recursively partitioning the data until each partition is pure" (col. 3:40-41). *Rastogi et al.* then describes two conditions for splitting the data: if the data A is numeric, then the split is of the form $A < v$, and if data A is categorical, then the split is of the form $A \in V$. Then, *Rastogi et al.* chooses the "split with the least entropy" (col. 4:38).

Nowhere does *Rastogi et al.* describe "cluster analysis" or even a split based on any type of cluster analysis. In fact, *Rastogi et al.* nowhere mentions a "cluster." The Office Action correctly acknowledges that *Rastogi et al.* does not explicitly teach cluster analysis "based on distances," and then relies on *Shimoji et al.* as disclosing "a method of clustering a set of data by using a clustering error based on distances between the data and respective one or more centers of the one or more clusters" (p. 6) *Shimoji et al.* is directed to clustering data based on entropical scheduling, where the assignment of a cluster to each data, for the update of the cluster center, is probabilistic, where the probabilities that each data belongs to individual clusters depend on the distances to the corresponding cluster centers (Abstract). Nowhere does *Shimoji et al.* disclose or suggest "performing a cluster analysis along the selected feature to group the data into one or more clusters based on distances between the data and respective one or more centers of the one or more clusters." In fact, the data of *Shimoji et al.* is defined over a d-dimensional space, and the clustering error is "measured by the Euclidean distance" in d-space, (Introduction, page 2423, right column) and thus there is no suggestion for a cluster analysis "along the selected feature."

As motivation for a combination of *Rastogi et al.* in view of *Shimoji et al.*, the Office Action contends, "to combine clustering error as taught by Shimoji to analyze a cluster when

grouping data into one or more cluster of a decision tree." However, the Office Action fails to explain how one skilled in the art would utilize the "clustering error" of *Shimoji et al.* (Equation (1), page 2423) in combination with *Rastogi et al.*, which nowhere even mentions "clusters," much less any "distances" between any data and other objects. In fact, even if *Rastogi et al.* had any clusters, any type of added "cluster analysis" would be technically infeasible, as *Rastogi et al.* already discloses an equation for entropy for a set of records, based on relative frequencies of respective classes in the set (e.g., "the more homogeneous a set is with respect to the classes of records in the set, the lower is the entropy"), and an equation for entropy of a split to divide the set, and states, "Consequently, the split with the least entropy best separates classes, and is thus chosen as the best split for a node." Thus, there is no motivation to combine *Rastogi et al.* and *Shimoji et al.*, other than impermissible hindsight. Thus, the rejection of claims 1-3 and 18-22 based on *Rastogi et al.* in view of *Shimoji et al.* should be withdrawn.

With regard to claims 10, 12, 16, 27, 29, and 33, the rejection over *Janikow* is also respectfully traversed because *Janikow* teaches against the proposed modification. The Office Action contends that it would have been obvious "to modify the Janikow method by using function f_2 as the membership function ... in order to split a node." However, *Janikow*, p. 9, teaches against just such a use: "To define the decision procedure, we must define f_0, f_1, f_2, f_3 for dealing with samples presented for classification. These operators may **differ from those used for tree-building**—let us denote them g_0, g_1, g_2, g_3." Thus, *Janikow* discloses a distinction between classification functions and tree building functions, and one of ordinary skill in the art would **not** be motivated to disregard *Janikow*'s distinctions and principle of operation when making modifications of its method.

With regard to claims 1-5 and 18-22, the rejection based on *Janikow* in view of *Choe et al.* is also traversed since the proposed modification of *Janikow* to use *Choe et al.*'s classification

system also ignores *Janikow*'s distinction between classification functions and tree building functions. Because of this distinction, *Janikow* actually teaches against using a function such as in *Choe et al.* for tree building (cf. claims 1 and 18: "constructing one or more arcs of the decision tree").

The anticipation rejection of claims 17 and 34-36 over *Background* is also respectfully traversed. The *Background* does not disclose the "selecting the one of the features corresponding to the maximal partition coefficient." In FID3, on the other hand, an attribute is chosen based on a maximum information gain, which is based on entropy instead of partition coefficients (*Background*, p. 4, line 13, cf. p. 3, line 17; *Janikow*, p. 7, col. 2). The Office Action appears to construe the recited "maximal partition coefficient" to read on a maximum information gain. The basis for this unusual interpretation appears to be a phrase in the specification that explains a property of the partition coefficient as "which quantifies the goodness of the clustering" as if anything that might have some connection to clustering must be a partition coefficient. But nothing in the Background asserts that the information gain has anything to do with goodness of clustering. A fuller discussion of information gain can be found in *Janikow*, but one of ordinary skill in the art would see that the information gain in *Janikow* is used on the tree-building side, **not** on the classification side (claim 17 recites "building the decision tree"). Thus, a person of ordinary skill in the art would not understand information gain to describe goodness of clustering.

Applicants submit that it is only by impermissible hindsight from the Applicants' disclosure of a "**unified approach** to extracting both the decision tree and the (crisp or fuzzy) clusters" that the prior art distinctions as exemplified in *Janikow*'s distinctions can be eroded both in proposing modifications thereto and to stretch the understanding of information gain to a point beyond which persons of skill in this art would accept.

The dependent claims are allowable for at least the same reasons as their independent claims and are individually patentable on their own merits. The additional secondary references, *Hall et al.* and *Shafer et al.* do not cure the above-described deficiencies in the applied art.

Therefore, the present application, as amended, overcomes the objections and rejections of record and is in condition for allowance. Favorable consideration is respectfully requested. If any unresolved issues remain, it is respectfully requested that the Examiner telephone the undersigned attorney at 703-425-8501 so that such issues may be resolved as expeditiously as possible.

Respectfully Submitted,

DITTHAVONG & CARLSON, P.C.

August 20, 2004
Date

Margo Livesay
Margo Livesay, Ph.D.
Reg. No. 41,946

Stephen C. Carlson
Reg. No. 39,929
Attorneys for Applicant(s)

10507 Braddock Rd
Suite A
Fairfax, VA 22032
Tel. 703-425-8501
Fax. 703-425-8518